Detection Anomaly of Network Datasets with Honeypots at Industrial Control System

Abbasgholi Pashaei[1], Mohammad Esmaeil Akbari[2*], Mina Zolfy Lighvan[3,] Asghar Charmin[4]
1,2,4- Department of Electrical Engineering, Ahar Branch, Islamic Azad University, Ahar, Iran
3- Department of Electrical and Computer Engineering Faculty, Tabriz University, Tabriz, Iran
Email: a-pashaei@iau-ahar.ac.ir, m-akbari@iau-ahar.ac.ir(Corresponding author), a_charmin@sut.ac.ir
Email: mzolfy@tabrizu.ac.ir

Receive Date: 10 March 2022, Revise Date: 15 April 2022, Accept Date: 10 June 2022

Abstract:Thedevelopment of ICS 4.0 industry-specific cybersecurity mechanisms can reduce the vulnerability of systems to fire, explosion, human accidents, environmentaldamage, and financial loss. Honeypots are computer systems that are deployed expressly to trick attackers into thinking they are real computers. Given that vulnerabilities are the points of penetration into industrial systems, and using these weaknesses, threats are organized, and intrusion into industrial systems occurs. As a result, to learn about an attacker's behavior, tactics, strategies, and signatures, the EIDS is used to collect information on cyber-attacks, proving it to be a more helpful tool than earlier traditional ways. Attacks collected by honeypot software expose the attackers' source IP addresses as well as the target host that became a victim of the assaults. This paper proposes a novel Honeypot enhanced industrial Early Intrusion Detection System (EIDS) using Machine Learning (ML). The performance of EIDS is evaluated with ML, and the experimental results show that the proposed EIDS detects anomalous behavior of the data with a high detection rate, low false positives, and better classification accuracy.

Keywords:Intrusion Detection System, Honeypot, Machine Learning, Anomaly Detection.

1. Introuduction

Security IDS management in Machine Learning-based Industrial LAN Networks Employing Honeypots systems are some industrial uses of SCADA networks [1], gas and oil flow control through pipes in the power plant industry [2, 3], output monitoring in power smart grid systems [4-7], monitoring products distribution in manufacturers [8-11], controlling railway and other transportation lines[12], and processing management in chemical areas[13]. Through the recent advancements in hardware technologies and multiple algorithms such as Artificial Intelligence (AI), ML, Deep Learning (DL), Data Mining (DM), radio communications, telemetry, and computer processing, almost all industries control processes remotely through SCADA.

As mentioned earlier, to satisfy the applications' needs, SCADA networks should monitor geographically distributed properties securely[14]. The earliest ICSs consisted of straightforward point-to-point networks that connected a monitoring device to out-of-the-way pieces of equipment. However, security requirements in these simple systems to support communications between the central monitoring and out-of-the-way equipment were not provided[15].

With the development of 4th generation industries in recent years, modern SCADA networks integrate with the smart sensors, Internet of Things (IoT), AI, cloud-based digital data stored systems, and Big data analytics[16]. Although the combination of emerging technologies could improve the infrastructure and maintenance costs, system performance, and interoperability, it was associated with new security challenges in the near real-time environments, including access control, classic Intrusion Detection Systems (IDSs), protocol vulnerability assessment, facilities, and operating systems (OSs) safety, key management in cryptography algorithms[17, 18] and crosstalk of communication equipment [19]. These cybersecurity attacks are becoming more sophisticated and carrying risks like an explosion in industrial environments, dangers to human life, and financial damage[20]. Thus, achieving a secure SCADA network for ICSs enhances industrial applications' security and performance of cyber-systems[21].

Honeypots industrial networks are responsible for attracting attackers, misleading them in the attack, and simulating basic industrial infrastructure. They also obtain the attacking device's characteristics, gain valuable information about the attacker, and identify the attack pattern. Therefore, imitating industrial control infrastructures protects industrial facilities' main sites against destruction and attack[22].

So far, some researchers have exploited artificial intelligence-based approaches to guarantee SCADA networks' security in ICS fields. Although the existing mechanisms enhance the performance of IDSs in industrial environments, they struggle with unacceptable performance. Besides, some of the IDS models only focus on the cybersecurity arena and ignore the process event states in physical ICS environments[23]. Furthermore, most ML-based ICS networks focus on cyberattack detection in industrial applications, and they do not describe the real impact of threats. Therefore, developing a new IDS with industrial Honeypot for networks to improve industrial infrastructure requirements security is necessary.

In this essay, part 2 reviews relevant works, section 3 examines the honeypot industrial's suggested approach, and section 4 studies the honeypot industrial's early intrusion detection system (EIDS) datasets. The discussion of the dataset and statistical findings is addressed in section 5 using a variety of datasets, and the conclusion is discussed in section 6 separately.

2. Related works

Securityisasignificantchallengetosatisfythere quirements of SCADA applications in ICS. A broad range of approacheshas been presented in the literature to address this issue. A summary of the previous woks is given in Tab.1.

Authors	Publication year	Summary
Pashaei et al. [24]	2022	They proposed a honeypot-assisted industrial control system to detect replication attacks on wireless sensor networks
Mashima et al. [25]	2017	Proposed an intelligent grid Honeypot system.
Dalamagkas et al. [26]	2019	Reviewed the Honeypot-based techniques in smart grids
Shi et al. [27]	2019	Proposed a dynamic property Honneypot based on Blockchain to distinguish between real and fake resources in the system.

Luo et al. [28]	2017	Proposed an IoT Honeypot to modify IoT security.
Nursetyo et al. [29]	2019	Proposed a Honeypot system to identify intruders by evaluating network server security techniques.
Bykara and Das [30]	2018	Honeypot was combined with IDS to increase the effectiveness of real-time intrusion detection.
ZiaieTabari and Ou [31]	2020	A multi-faceted and multi-phase IoT Honeypot ecosystem was designed to obtain information from cyberattackers and examine them in the IoT systems.
Yang et al. [32]	2019	Incorporated DL networks with SCADA-based systems to protect ICSs from conventional and network-based cyberattacks. Used CNNs to automatically extract salient features and took benefit from a re-training mechanism to improve the performance on new attacks. The proposed DL-based framework improved the detection accuracy and identified advanced-emerged threats.
Gao et al. [33]	2020	Used a feedforward neural network and an LSTM to develop a DL-based IDS able to detect temporally correlated and uncorrelated attacks in SCADA-based systems.
Perez et al. [34]	2018	Used SVM and RF for detecting intrusions not seen in the database and concluded that the RF outperformed the SVM in providing the security of the SCADA systems in ICSs.
Sheng et al. [35]	2021	Introduced a cyber-physical identification plan for evaluating risk levels of intrusions against vulnerable industrial systems with deficiencies in control devices and protocols to encounter threats. Communication patterns and states of devices were extracted the characterize the system structure. Any violation of then plan was considered as a false or network-based cyberattack.
Khan et al. [36]	2019	Presented a hybrid, multi-level method for intrusion detection in SCADA networks to deal with unbalanced data in ICSs. The method employed a KNN rule plan to improve the accuracy of detection. Although the technique focused on the cybersecurity arena industrial, it ignored the process states in industrial applications' physical environments.
Qian et al. [37]	2020	A secure mechanism for detecting cyber and physical aggression in SCADA networks is was introduced to tackle physical field challenges and detect processing attacks such as the Man-in-the-Middle (MITM). The mechanism also proposed a Nonparallel Hyperplane-based Fuzzy (NHF) classifier for dataset classification. The comparisons proved that this hybrid mechanism's performance was preferable to the parallel hyperplane of the SVM in the cyber field.
Bulle et al. [38]	2020	A reliable host-based IDS through the OS diversity has been introduced to detect new kinds of threats in SCADA networks. SCADA communications over time were evaluated in an ICS to select the most reliable OS in the system. Experiments showed that choosing the most suitable OS enhances IDS accuracy compared to the single operational system-based environments.

The research in [39] concentrated on critical issues concerning Internet of Things (IoT) technology. They created a honeypot using reinforcement learning (RL) to detect attacks caused by DDoS and Man in the Middle attacks. They discovered that a honeypot built using reinforcement learning can detect up to 99.96% of attacks and outperform previous honeypots in terms of performance.

According to the literature, the combined IDS-based methods improvedeffectiveness in SCADA networks; However,ML-basedapproaches

in the literature have focusedoncyberattack detection against SCADA networks in industrialenvironments,andtheyhavenotcoveredtherealimpactof threatsonICS. Consequently, it is necessary to study the effect of incorporating different deep and shallow machine learning algorithms with intrusion detection systems. The contribution of this paper is to investigate the performance of a Honeypot system combined with different machine learning and deep learning system to enhance its accuracy and computational speed.

3. Proposed Methodology

A. Research Model

The investigated strategy in this work is considered as follows:

- The issue undermining arrange security was recognized; the framework shortcomings of conventional IDS were evaluated from previous paper works performed and then added to the new system topology's general plan.
- The framework of topology for intrusion discovery testing and ongoing evaluation was designed. And average rankings for all forecasting techniques are calculated by:

$$R_i = \frac{1}{n}\sum_{i=1}^{n} r_i^{\,j} \qquad (1)$$

Where each pair of i is given a rank. Ranks are represented by r_i^j the notation $\left(1 \le j \le k\right)$ and range from 1 (least mistake) to k (greatest error).

- Attack area zones were decreased by dividing networks into logical sections and restricting host-to-network communications direction.
- Blacklists and whitelists were used to protect topologies and architectures designed against potentially harmful applications.
- After implementing and testing the framework, the assessment was performed simultaneously as in previous experiments. The one-hot encoding is used in this instance. Additionally, the Min-Max skill is then applied to limit the range of the encoded data to [0,1].

$$x^{*} = \frac{x - \min}{\max - \min} \bullet \qquad (2)$$

The method proceeded until the aiming results were accomplished.Access logs were analyzed, and anomalies with ML were verified.andThe maximum Softmaxprobability of the output cells determines the classification category.

$$S_i = \frac{e^{v_i}}{\sum_{i=1}^{n} e^{v_i}} \bullet \qquad (3) \bullet$$

In the ultimate operation stages, the interruption discovery framework was re-experimented step by step to guarantee proper operation.A complete model of the EIDS algorithm is shown in Fig. 1, based on the previous points regarding the investigation strategy. The experiment's schematic and investigation strategy demonstrate that the chosen EIDS framework must coordinate the requests of the EIDS from the starting of the determination life cycle to the assessment alteration of the framework.

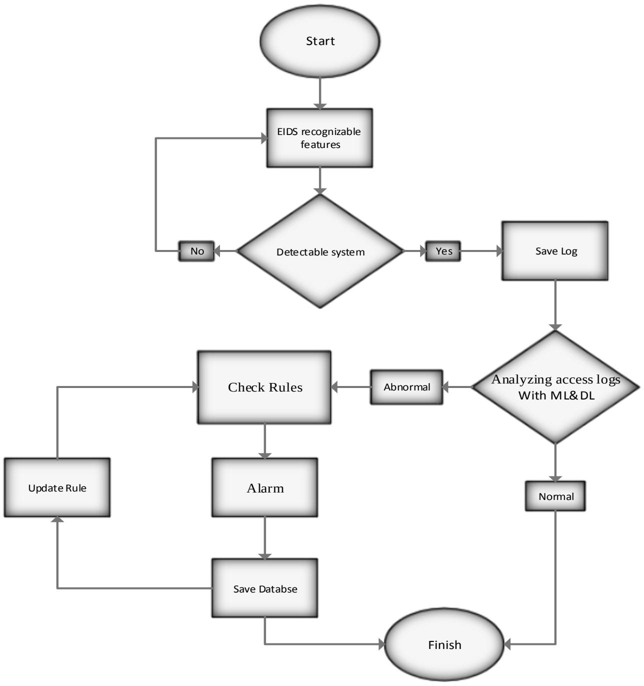


Fig. 1. The EIDS algorithm life cycle.

Fig. 2, Generalizes process-based computer automation architecture to industrial control systems. Standard options for implementing computer-based industrial control processes are servers or computers, PLC, RTU, etc. All are interfaced through input/output subsystems for processing equipment (e.g., sensors and valves). In addition, PLCs typically have access to other computers that support industrial facility operations over the LAN and wireless.

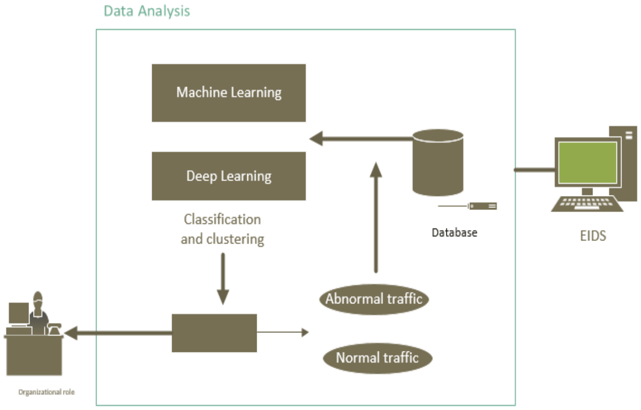


Fig. 2. The proposed architecture for data analysis with ML on EIDS logs.

This research uses the proposed model and architecture for intrusion detection Honeypot to help ML. The proposed methods to be used in behavioral experiences Honeypot has also been illustrated.

4. Datasets In Honeypot Industrial Early Intrusion Detection System (EIDS)

The Honeypot early detection system's success depends on the correct choice of factors and features used in tracking attacks. This paper presents Honeypot EIDS technology using DL and SL algorithms because DL and SL technologies are suitable for identifying attackers by extracting and collecting features using attackers' performance logs. Prior to entering this section's topics, it is necessary to state some items to help analyze the results obtained briefly. To execute the code and analyze the results, Python Anaconda, Jupiter Notebook distribution will be used. The notebook client allows extensive, scalable, and reproducible use of code. The new algorithm is disregarded and the next one is tested in its place if there isn't a ratio to raise the set's accuracy. The decision to begin with the poorest models is justified by the way that the ratio of the first models gradually drops when more are added.

$$p = \sum_{i=1}^{N} \lambda_i \, p_i \qquad (4)$$

This technique produces ensembles, each specialized in a certain type of Dataset. The prediction p from each ensemble is the weighted sum of the pi predictions from each algorithm i by a ratio λ_i.

Jupiter Notebook is an open platform in the browser environment for prototyping and researching analysis. After installing the necessary ML libraries, such as Pandas and SciKit Learn, and reciprocal libraries for DL, tasks for each project will be created in separate environments.

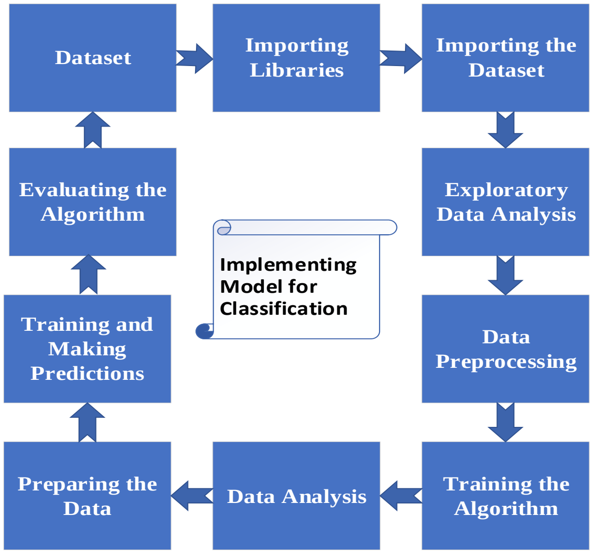


Fig. 3. The implementation model for classification.

The accepted data should be easily obtained for the proposed EIDS and reflect the host or network's behavior. Consider that building a dataset is a complex and time-consuming process. Therefore, using a benchmark dataset helps to facilitate the diagnosis time. Because the benchmark data sets are valid, they produce and extract the experimental results in the laboratory research more convincing and allow the results in the proposed method to be compared with previous studies. To extract the most optimal and efficient detection model for the stored data from the Honeypot, EIDS logs are used in the laboratory for this research and ensured its results and accuracy. Datasets is used: EIDS dataset explained in the following sections.

Therefore, an executable implementation model is designed to classify the datasets mentioned according to Fig. 3, which can be run and done with this method, such as importing the dataset, data preprocessing, data analysis, etc.

A. EIDS Database

As the logs are generated in the industrial network, IDS Snort completes the dataset of the EIDS. The proposed EIDS database is a lightweight and potent tool that authorizes the system to detect intrusion of malicious network traffic early. So, almost any threat that crosses the network can be identified by defining flexible and robust rules. To provide the mentioned needs, a solution to process the alert data of this huge dataset is needed.

Therefore, the CSV format for processing alert data is used, which is the most flexible and compatible method for data collection. To configure IDS Snort to use the CSV output format, add the following command to the Snort. conf file:

```
output alert_csv: alert.csv default
```

This command configures IDS Snort to create a CSV log file called alert.csv in the configuration log using the default output, and 30 features can be extracted from IDS Snort in the following as Tab. 2.

Tab. 2. Generated features for the EIDS database.

Feature	Feature	Feature
time	icmpseq	icmpid
icmpcode	date	sig_generator
icmptype	iplen	dgmlen
id	tos	ttl
tcpwindow	tcpln	tcpack
tcpseq	tcpflags	ethlen
ethdst	ethsrc	dstport
dst	srcport	src
proto	msg	sig_rev
sig_id	timestamp	

Honeypot EIDSs are used to detect cyberattacks in a network of the ICS. Thus, various studies have been conducted on high-performance datasets based on ML techniques. In the field of IDS, famous datasets are available for evaluations like NSL-KDD intrusion detection datasets, CIC-IDS 2017, and Kyoto 2006 datasets. However, these datasets do not reflect recent cyberattack trends in the proposed research. For this reason, the EIDS dataset from the same traffic data from the study with the latest Snort Log is refined. Besides, the new dataset is evaluated by applying several ML techniques and comparing the datasets' classification results.

1)False-Negative Rate (FNR)

According to the specified type of action in equation 5, FNR means when the sensor detects healthy traffic as malicious traffic and acts on this traffic. According to the applied signatures on the system, the proposed healthy traffic will be blocked, and it will not be allowed to pass, or if there are actions for logging, it will generate logs and alerts. Therefore, it would be difficult for the system administrator to root the reasons for the created FNs when checking logs and alerts. In general, having many FNs in the network hurts network performance and must be identified, investigated, and managed.

$$\text{FNR} = \frac{FN}{TP + FN} \tag{5}$$

2)False-Positive Rate (FPR)

The FPR in equation 6 means when the sensor does not detect malicious traffic. In this case, the proposed network is endangered because malicious traffic passes through the proposed network without being detected and blocked and can damage the network resources. This non-detection of malicious traffic can be due to various reasons. For example, the sensor signatures have not been updated, and new signatures have not been received, or the sensor settings have not been done correctly. Therefore, the sensor has not been able to function correctly, or this malicious traffic is a new method that has not yet been addressed.

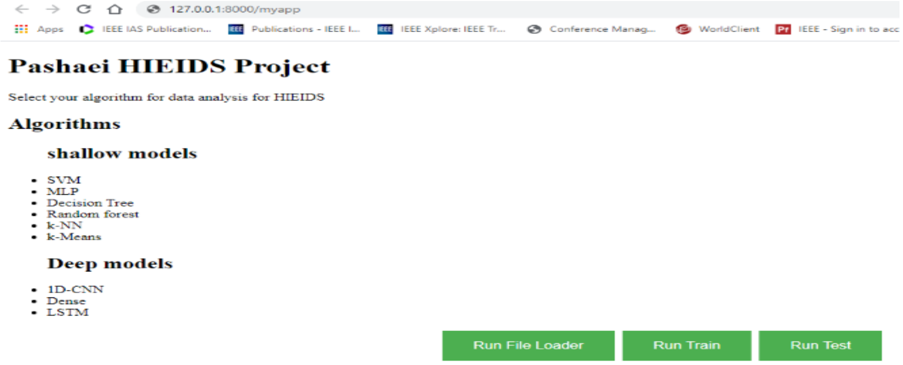
$$\text{FPR} = \frac{FP}{FP + TP} \tag{6}$$

The area under the chart is represented by the AUC (ROC). The end performance of the category will be more ideal the more of this category there are in the category. The ROC chart can be used to evaluate how well each category is performing. The AUC index calculation depends on (28)

$$\begin{aligned} &1 \to TPR \to y(axis) \\ &2 \to FPR \to x(axis) \\ &1,2 \to AUC = \int_0^1 TPR\left(FPR^{-1}(x)\right)dx \end{aligned} \tag{7}$$

B. Monitoring and data mining application

This program uses 1056 lines of code and some other codes, such as web recall applications, algorithm connectors, etc., to execute the learning model algorithms in the EIDS project program. As shown in Fig. 4, the design is done in a convenient, simple, and user-friendly way for learning model systems, so that the steps of uploading CSV files can be done directly from the EIDS system log storage, and real-time analysis can be done to detect new attacks. The advantage of this is that it increases the percentage of reliability and reliability in detecting the early intrusion system alongside the detection system and gives us a deeper and more comprehensive understanding of the detection and investigation of various attacks to analyze the behavior and future actions of attackers.



Pashaei HIEIDS Project

Select your algorithm for data analysis for HIEIDS

Algorithms

shallow models

- SVM
- MLP
- Decision Tree
- Random forest
- k-NN
- k-Means

Deep models

- 1D-CNN
- Dense
- LSTM

Fig. 4. Schematic of a program designed to detect, analyze, and perform EIDS ML models .

5. Discusion on the Daaset and Statical Observations

The logs of all incoming and outgoing traffic that interacted with the Honeypot sensors in any way are stored in the MySQL database of the Linux-based OS. As mentioned earlier, the EIDS dataset has been introduced for the accuracy of the operation and the ability to detect the intrusion of logs stored in the EIDS database. Being compared with standard evaluation methods is essential, which makes the evaluation results reliable for this dataset.

Therefore, the four databases have been analyzed and processed separately. The results of separate analyses of each item according to the measurements are obtained and shown in the form of tables and diagrams in this section with explanations. Explanation about these analyzes, processes, graphs, and results were performed with a dedicated program written in Python software for this study.

For the NSL-KDD dataset, processing and analyzing the obtained results from the designed program is given in Tab.3. Tab. 3 calculated obtained results for detecting anomalies traffic for the algorithms used in the research for the NSL-KDD dataset. The obtained results from 7 algorithms in Tab.3 are shown in Fig. 5 as a bar chart. In Fig. 8(a), the results of accuracy obtained from Tab. 3 are shown for seven algorithms. In Tab. 3, two essential criteria, accuracy and F1-Score, are calculated from 7 algorithms. As stated, the accuracy criteria demonstrates that the LSTM algorithm outperforms competing techniques.

Tab. 3. Obtained results for detecting anomalies traffic for the algorithms used in the research for the NSL-KDD dataset.

Method	Accuracy	Recall	Precision	F1
Tree	0.763	0.662	0.895	0.761
KNN	0.772	0.652	0.924	0.765
MLP	0.797	0.696	0.929	0.796
SVM	0.763	0.639	0.920	0.754
Dense	0.815	0.706	0.959	0.813
CNN1D	0.770	0.652	0.922	0.764
LSTM	0.886	0.976	0.847	0.907

Another criterion is the F1-Score criterion, which is a combination of the R and P criteria, and again, the LSTM algorithm works better, as can be seen in Fig. 5. the dense algorithm in Fig. 5 offers a very high P, but it doesn't mean that this algorithm has high accuracy and F1-Score as well.

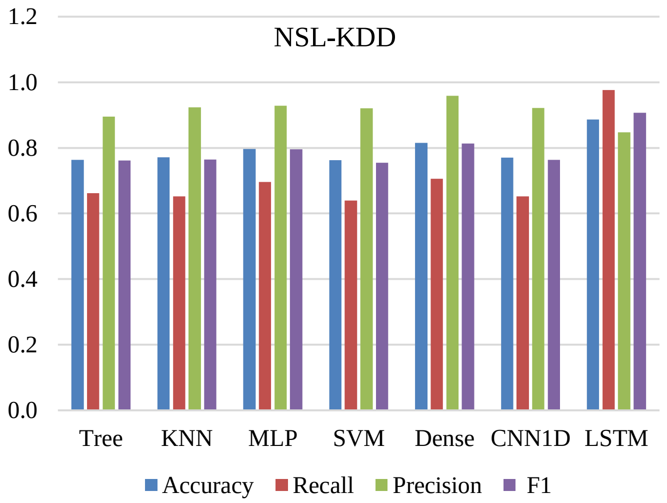


Fig. 5. Measured accuracy, R, P, and F1-Score to detect traffic anomalies in the algorithms used for the NSL-KDD dataset using Python program.

Similarly, Tab. 4 shows the obtained results from the CIC-IDS2017 dataset. The obtained results from 7 algorithms are demonstrated in Fig. 6 and 8(b). In Fig.6, all seven algorithms are shown in a bar chart, while in Fig.8(b), only the accuracy is evaluated. As described in the accuracy and F1-Score criterion, the KNN Shallow Learning algorithm performs better than other methods, as shown in Fig. 6. The DT algorithm in Fig. 6 offers a very high P, but it doesn't mean that this algorithm has a high F1-Score.

Tab. 4. Obtained results for detecting anomalies traffic for the algorithms used in the research for the CIC-IDS2017 dataset.

Method	**Accuracy**	**Recall**	**Precision**	**F1**
Tree	0.998	0.852	0.995	0.918
KNN	1.000	0.987	0.985	0.986
MLP	0.997	0.958	0.835	0.892
SVM	0.994	0.854	0.736	0.791
Dense	0.997	0.976	0.836	0.900
CNN1D	0.995	0.863	0.808	0.808
LSTM	0.680	0.912	0.037	0.070

Next, Tab. 5 shows the obtained results from the Kyoto 2006 dataset. The obtained results from 7 algorithms are demonstrated in Fig. 7 and 8(c). In Fig.7, all seven algorithms are shown in a bar chart, while in Fig.8(c), only the accuracy is evaluated. As described in the accuracy and F1-Score criterion, the DT Shallow Learning algorithm performs better than other methods, as shown in Fig. 7. The SVM algorithm in Fig. 7 offers a high R, but it doesn't mean that this algorithm has a high F1-Score.

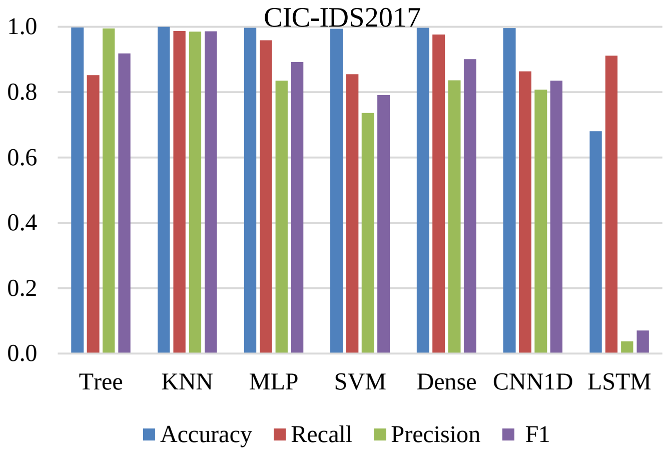


Fig. 6. Measured accuracy, R, P, and F1-Score to detect traffic anomalies for the CIC-IDS2017 dataset using Python program.

Tab. 5. Obtained results for detecting anomalies traffic for the algorithms used in the research for the Kyoto 2006 dataset.

Method	Accuracy	Recall	Precision	F1
Tree	0.9995	0.9995	0.9993	0.9994
KNN	0.9901	0.9993	0.9775	0.9883
MLP	0.9743	0.9806	0.9586	0.9695
SVM	0.8774	0.9881	0.7778	0.8704
Dense	0.8521	0.6455	0.9994	0.7844
CNN1D	0.9794	0.9514	0.9992	0.9747
LSTM	0.6262	0.5344	0.5533	0.5436

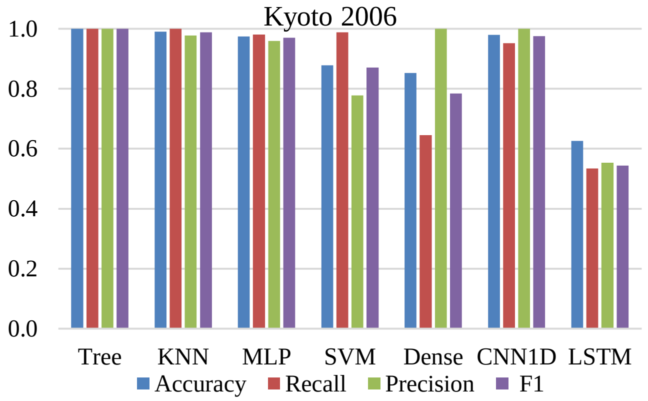


Fig. 7. Measured accuracy, R, P, and F1-Score to detect traffic anomalies for the Kyoto 2006 dataset using the Python program.

Features for the EIDS database considered to be maximally effective features that will help the data stored as logs to be used in the best possible way to detect anomalies in the EIDS system. A pair plot in Fig.9 is shown to prove this subject in EIDS systems. A pair plot is a distribution diagram that basically draws a common diagram for all possible combinations of numeric and Boolean columns in the EIDS database and sends the EIDS data frame as a parameter to the pair plot function. All null values were removed from the data before the pair plot command was executed. Common diagrams of all numeric and Boolean columns in the EIDS database can be viewed in the output of the pair diagram. The batch column name is given to the hue parameter to add categorical column information to a pair chart. For example, to draw label information on a pair chart, information about normal logs in blue and information about abnormal logs (attack logs) in orange are visible in the output (as shown in the descriptions and abbreviations). This is clearly seen in the common diagram at the top left that most early detection of right logs is related to attacks.

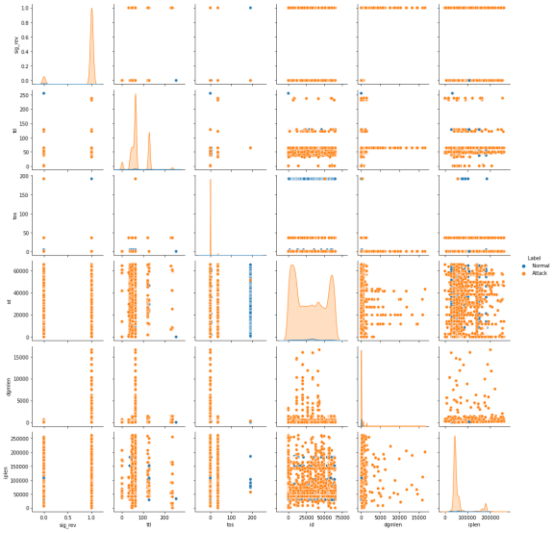
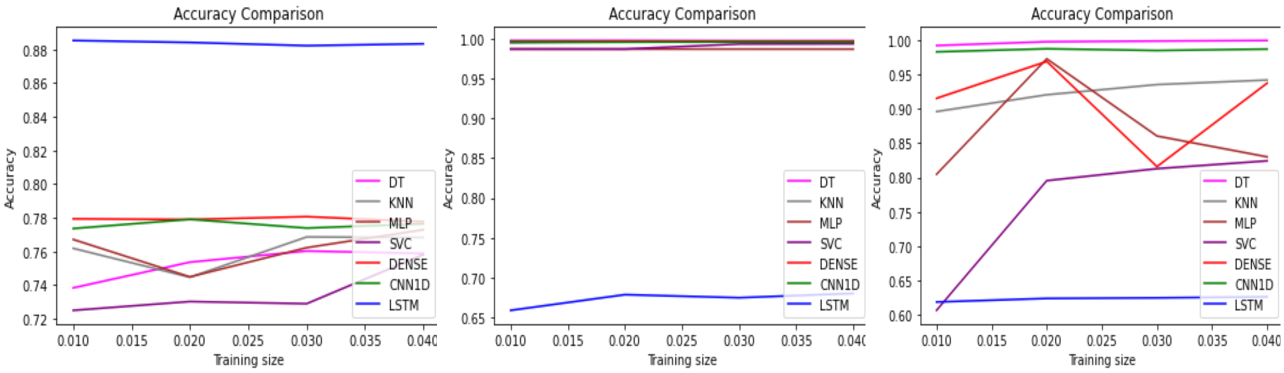


Fig. 9. Measured sns. pair plot (df, hue = 'Label')
for detecting traffic anomalies with Python simulation for the EIDS database.



(a) NSL-KDD (b) CIC-IDS2017 (c) Kyoto 2006

Fig. 8. Measured accuracy to detect traffic anomalies using Python simulation program for the algorithms used in the NSL-KDD, CIC-IDS2017, and Kyoto 2006 dataset.

Finally, for the model presented in this study called EIDS, Tab. 6 shows the obtained results for the EIDS database. The obtained results from 7 algorithms are demonstrated in Fig. 10 and11. The accuracy criterion shows that the DT, KNN, MLP, and SVM Shallow Learning algorithm and dense layer and CNN1D from Deep Learning algorithms perform better than other methods shown in Fig. 10.

Tab. 6. Obtained Results for detecting traffic anomalies

for the algorithms used in the research for the EIDS database.

Method	Accuracy	Recall	Precision	F1
Tree	0.9995	0.9995	0.9993	0.9994
KNN	0.9901	0.9993	0.9775	0.9883
MLP	0.9743	0.9806	0.9586	0.9695
SVM	0.8774	0.9881	0.7778	0.8704
Dense	0.8521	0.6455	0.9994	0.7844
CNN1D	0.9794	0.9514	0.9992	0.9747
LSTM	0.6262	0.5344	0.5533	0.5436

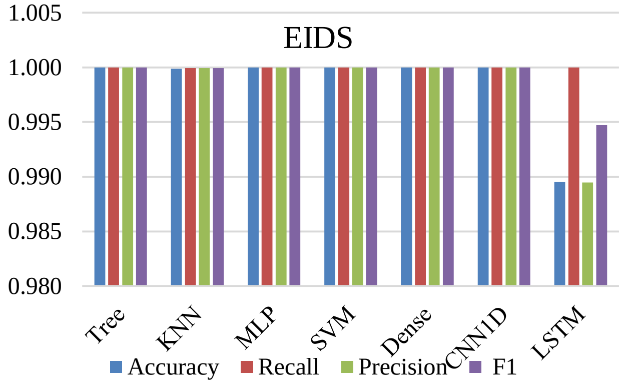


Fig. 10. Measured accuracy, R, P, and F1-Score to detect traffic anomalies for the EIDS database using Python program.

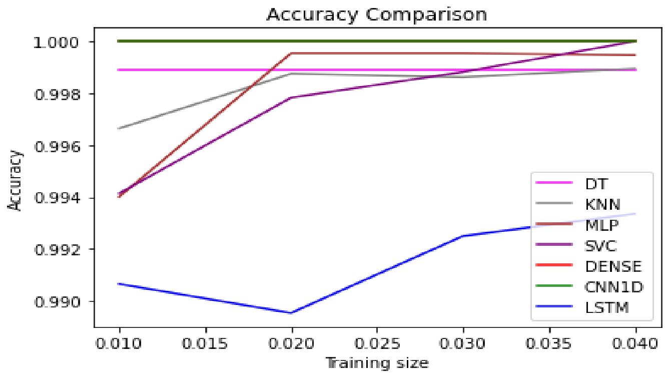


Fig. 11. Measured accuracy in detecting traffic anomalies with Python simulation program for the algorithms used in the EIDS database.

Conclusion

Various datasets, such as NSL-KDD, CIC-IDS2017, and Kyoto 2006, were used to implement a comprehensive plan for the classification of industrial networks, and a database was created based on the best features. Finally, the accuracy index was evaluated in a fully equipped ICS laboratory for the three reference datasets and a database in the proposed method (EIDS) in ML. The accuracy of EIDS has increased compared to the threementioned datasets. The accuracy of EIDS on the main database has increased by 31% in test data DT, 29.59% in test data KNN, 25.50% in test data MLP, 31.06% in test data SVM, 22.66% in Dense layer test data, 29.80% in test data CNN1D, and 66% in test data LSTM compared to NSL-KDD. The accuracy of EIDS on the main database has increased by 0.20% in test data DT, 0.20% in test data KNN, 0.31% in test data MLP, 0.60% in test data SVM, 0.29% in Dense layer test data, 0.46% in test data CNN1D, and 45.61% in test data LSTM compared to CIC-IDS2017. The accuracy of EIDS on the main database has increased by 0.05% in test data DT, 0.99% in test data KNN, 2.64% in test data MLP, 13.97% in test data SVM, 17.35% in Dense layer test data, 2.10% in test data CNN1D, and 58.03% in test data LSTM compared to Kyoto 2006.

According to the obtained results, the program developed for this research significantly improved the analysis of the EIDS database for early intrusion detection compared to other datasets. The performed design with high accuracy can detect abnormal traffic in industrial facilities by its expanded sensors in the network of industrial facilities. The proposed EIDS design works well in industrial environments. Therefore, it is an efficient and integrated system for cybersecurity to counter future attacks and Zero days in industrial facilities.

References

[1] Pashaei, A., Akbari, M. E., Lighvan, M. Z., & Charmin, A. " Machine Learning-based Industrial LAN Networks Using Honeypots". Journal of Artificial Intelligence in Electrical Engineering, 2023.

[2] Lu, H., Iseley, T., Behbahani, S., & Fu, L. (2020). Leakage detection techniques for oil and gas pipelines: State-of-the-art. Tunnelling and Underground Space Technology, 98, 103249.

[3] Kim, S., Heo, G., Zio, E., Shin, J., & Song, J. G. (2020). Cyber attack taxonomy for digital environment in nuclear power plants. Nuclear Engineering and Technology, 52(5), 995-1001.

[4] Kermani, M., Adelmanesh, B., Shirdare, E., Sima, C. A., Carnì, D. L., & Martirano, L. (2021). Intelligent energy management based on SCADA system in a real Microgrid for smart building applications. Renewable Energy, 171, 1115-1127.

[5] Wang, X., Zhao, Q., Yang, X., & Zeng, B. (2020). Condition monitoring of wind turbines based on analysis of temperature-related parameters in supervisory control and data acquisition data. Measurement and Control, 53(1-2), 164-180.

[6] Kermani, M., Parise, G., Shirdare, E., & Martirano, L. (2020, June). Transactive Energy Solution in a Port's Microgrid based on Blockchain Technology. In 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe) (pp. 1-6). IEEE.

[7] Wertani, H., Salem, J. B., & Lakhoua, M. N. (2020). Analysis and supervision of a smart grid system with a systemic tool. The Electricity Journal, 33(6), 106784.

[8] Aghenta, L. O., & Iqbal, M. T. (2019, May). Development of an IoT Based Open Source SCADA System for PV System Monitoring. In

2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE) (pp. 1-4). IEEE.

[9] Kermani, M., Carnì, D. L., Rotondo, S., Paolillo, A., Manzo, F., & Martirano, L. (2020). A nearly zero-energy microgrid testbed laboratory: Centralized control strategy based on scada system. Energies, 13(8), 2106.

[10] Pliatsios, D., Sarigiannidis, P., Lagkas, T., & Sarigiannidis, A. G. (2020). A survey on SCADA systems: secure protocols, incidents, threats and tactics. IEEE Communications Surveys & Tutorials, 22(3), 1942-1976.

[11] M. Kermani, E. Shirdare, A. Najafi, B. Adelmanesh, D. L. Carni and L. Martirano, (2021) "Optimal Self-scheduling of a real Energy Hub considering local DG units and Demand Response under Uncertainties," in *IEEE Transactions on Industry Applications*.

[12] Yu, S., Chang, H., & Wang, H. (2020). Design of Cloud Computing and Microservice-Based Urban Rail Transit Integrated Supervisory Control System Plus. Urban Rail Transit, 6(4), 187-204.

[13] Kosturko, J., Schlieber, E., Futch, S., & Nielson, S. (2018, October). Cracking a Continuous Flow Reactor: A Vulnerability Assessment for Chemical Additive Manufacturing Devices. In 2018 IEEE International Symposium on Technologies for Homeland Security (HST) (pp. 1-6). IEEE.

[14] Ghosh, S., & Sampalli, S. (2019). A survey of security in SCADA networks: Current issues and future challenges. IEEE Access, 7, 135812-135831.

[15] Bichmou, A., Chiocca, J., Hernandez, L., Hoffmann, R. W., Horsham, B., Lam, H., ... & Bibyk, S. (2019, July). Physical Cyber-Security of SCADA Systems. In 2019 IEEE National Aerospace and Electronics Conference (NAECON) (pp. 243-248). IEEE.

[16] Pashaei, A., Akbari, M. E., Lighvan, M. Z., & Charmin, A. Early Intrusion Detection System using honeypot for industrial control networks. Results in Engineering, 100576. (2022).

[17] Rosa, L., Freitas, M., Mazo, S., Monteiro, E., Cruz, T., & Simões, P. (2019). A comprehensive security analysis of a scada protocol: From OSINT to Mitigation. IEEE Access, 7, 42156-42168.

[18] Abou el Kalam, A. (2021). Securing SCADA and critical industrial systems: From needs to security mechanisms. International Journal of Critical Infrastructure Protection, 32, 100394.

[19] Pashaei, A., Andalib, A., & Banaei, H. A. (2014). Decrease of crosstalk phenomenon optic two-channel demultiplexer using resonant line defect cavity in 2D photonic crystal. Majlesi journal of Telecommunication devices, 3(1).

[20] Ferrag, M. A., Babaghayou, M., & Yazici, M. A. (2020). Cyber security for fog-based smart grid SCADA systems: Solutions and challenges. Journal of Information Security and Applications, 52, 102500.

[21] Ahn, S., Lee, T., & Kim, K. (2019, October). A Study on Improving Security of ICS through Honeypot and ARP Spoofing. In 2019 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 964-967). IEEE.

[22] Dutta, N., Jadav, N., Dutiya, N., & Joshi, D. (2020). Using honeypots for ICS threats evaluation. In Recent developments on industrial control systems resilience (pp. 175-196). Springer, Cham.

[23] Sun, Y., Tian, Z., Li, M., Su, S., Du, X., & Guizani, M. (2020). Honeypot Identification in Softwarized Industrial Cyber–Physical Systems. IEEE Transactions on Industrial Informatics, 17(8), 5542-5551.

[24] A. Pashaei, M. E. Akbari, M. Zolfy Lighvan, and A. Charmin, “A Honeypot-assisted Industrial Control System to Detect Replication Attacks on Wireless Sensor Networks”, Majlesi Journal of Telecommunication Devices, Vol. 11, No. 3, pp. 155-160, 2022.

[25] Mashima, D., Chen, B., Gunathilaka, P., & Tjiong, E. L. (2017, October). Towards a grid-wide, high-fidelity electrical substation honeynet. In 2017 IEEE International Conference on Smart Grid Communications (SmartGridComm) (pp. 89-95). IEEE.

[26] Dalamagkas, C., Sarigiannidis, P., Ioannidis, D., Iturbe, E., Nikolis, O., Ramos, F., ... & Tzovaras, D. (2019, June). A survey on honeypots, honeynets and their applications on smart grid. In 2019 IEEE Conference on Network Softwarization (NetSoft) (pp. 93-100). IEEE.

[27] Shi, L., Li, Y., Liu, T., Liu, J., Shan, B., & Chen, H. (2019). Dynamic distributed honeypot based on blockchain. IEEE Access, 7, 72234-72246.

[28] Luo, T., Xu, Z., Jin, X., Jia, Y., & Ouyang, X. (2017). Iotcandyjar: Towards an intelligent-interaction honeypot for iot devices. Black Hat, 1-11.

[29] Nursetyo, A., Rachmawanto, E. H., & Sari, C. A. (2019, October). Website and network security techniques against brute force attacks using honeypot. In 2019 Fourth International Conference on Informatics and Computing (ICIC) (pp. 1-6). IEEE.

[30] Baykara, M., & Das, R. (2018). A novel honeypot based security approach for real-time intrusion detection and prevention systems. Journal of Information Security and Applications, 41, 103-116.

[31] Ziaie Tabari, A., & Ou, X. (2020, October). A Multi-phased Multi-faceted IoT Honeypot Ecosystem. In Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (pp. 2121-2123).

[32] Yang, H., Cheng, L., & Chuah, M. C. (2019, June). Deep-learning-based network intrusion detection for SCADA systems. In 2019 IEEE Conference on Communications and Network Security (CNS) (pp. 1-7). IEEE.

[33] Gao, J., Gan, L., Buschendorf, F., Zhang, L., Liu, H., Li, P., ... & Lu, T. (2020). Omni SCADA intrusion detection using deep learning algorithms. IEEE Internet of Things Journal, 8(2), 951-961.

[34] Perez, R. L., Adamsky, F., Soua, R., & Engel, T. (2018, August). Machine learning for reliable network attack detection in SCADA systems. In 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) (pp. 633-638). IEEE.

[35] Sheng, C., Yao, Y., Fu, Q., & Yang, W. (2021). A cyber-physical model for SCADA system and its intrusion detection. Computer Networks, 185, 107677.

[36] Khan, I. A., Pi, D., Khan, Z. U., Hussain, Y., & Nawaz, A. (2019). HML-IDS: A hybrid-multilevel anomaly prediction approach for intrusion detection in SCADA systems. IEEE Access, 7, 89507-89521.

[37] Qian, J., Du, X., Chen, B., Qu, B., Zeng, K., & Liu, J. (2020). Cyber-physical integrated intrusion detection scheme in SCADA system of process manufacturing industry. IEEE Access, 8, 147471-147481.

[38] Bulle, B. B., Santin, A. O., Viegas, E. K., & dos Santos, R. R. (2020, October). A Host-based Intrusion Detection Model Based on OS Diversity for SCADA. In IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society (pp. 691-696). IEEE.

[39] A. Pashaei, M. E. Akbari, M. Zolfy Lighvan, and A. Charmin, “Honeypot Intrusion Detection System using an Adversarial Reinforcement Learning for Industrial Control Networks”, Majlesi Journal of Telecommunication Devices, Vol. 12, No. 1, 2023.